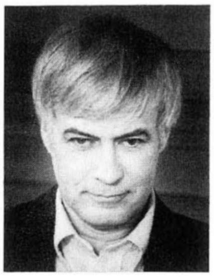
Editorial

The prospects for artificial intelligence

Robert K. LINDSAY
University of Michigan, Mental Health Research Institute, Ann Arbor, MI 48109, USA



Robert Kendall Lindsay was educated at Carnegie-Mellon University (B.S. in physics, 1956) and Columbia University (M.A. in psychology, 1957), and received his Ph.D. from Carnegie-Mellon (administration, 1961). His doctoral work was under the direction of Herbert A. Simon and Allen Newell, in a program that subsequently evolved into the first computer science department in the country. His dissertation, 'Toward the development of Machines that Comprehend Natural Language' was the first heuristically based natural language understanding program. During graduate school, he was the recipient of a Eugene Higgins Fellowship (Columbia), Ford Foundation Fellowships (Carnegie-Mellon), and a Social Science Research Council Fellowship (Carnegie-Mellon). Later, he spent a year at Stanford University on a National Science Foundation Science Faculty Fellowship studying logic, computer science, mathematics and philosophy.

Lindsay has held faculty appointments in psychology at The University of Texas at Austin and the University of Michigan. He has held visiting faculty appointments in computer science at Stanford University and the University of California at Irvine. He has lectured on artificial intelligence and heuristic programming at a number of places, including the University of California at Los Angeles, a Nato Advanced Study Institute, and a RAND Corporation Summer Institute. He has held research positions at the IBM Corporation Research Center and the New York State Biomedical Research Center. He has been a research consultant to the RAND Corporation, System Development Corporation, General Electric Company, and Xerox Corporation. At present he is Research Scientist at the University of Michigan Mental Health Research Institute.

Planning for the continuing and accelerating impact of the computer revolution is becoming more difficult as the field of artificial intelligence makes broader inroads into the automation of intellectual functions. Increasingly, decision making formerly done exclusively by humans will fall within the scope of automation, and the societal changes will be great. However, AI still lacks deep insight into the fundamental nature of human thought and experience, and theoretical breakthroughs of major proportion will be needed before machines will (or ought to) replace humans for many purposes. Surprisingly, many areas of human activity that appear to be purely logical and factual, and even seemingly trivial, will not soon be achievable by computers.

Keywords: Artificial intelligence, artificial understanding, technological forecasting, robotics.

As predictors of our technological future, scientists often fare no better or worse than movie makers or novelists. We are all aware of the oversights: the failures of vision, the unpredicted breakthroughs. At times, however, and perhaps contrary to the popular wisdom, scientists have promised more than they have delivered. In the case of artificial intelligence this has been so. Twenty five years ago some were predicting that remarkable feats of human intelligence would shortly be duplicated and even exceeded by artifacts. These predictions were based on what we now know were quite modest extrapolations of available computer power. The scientific community was *not* predicting the astonishing hardware progress in increased speed, decreased size and energy consumption, and lowered costs that we have in fact seen. It is remarkable in hindsight to realize that even the *modest* AI achievements of today would not have been possible were it not for these *unpredicted* increases in computing power.

The failure of these early forecasts has of course not diminished the flow of predictions, nor should it. We still must plan for the future, and that means going with our best guesses, fallible though they are. But which guesses are the best? The top level manager of today is confronted on all sides by claims about the impact that computer technology will have on all levels of business, government and society. Having observed the sweeping changes of the past two decades, he is apt to take these claims quite seriously. Will much of the work of his company, his own decision making included,

North-Holland
Human Systems Management 5 (1985) 95–98

be accomplished by 'expert systems'? Will the assembly lines be fully automated? Will labor costs be drastically reduced? Will new 'intelligent' devices flood the market and replace the company's current product line? Will unemployment lead to serious labor–management disputes? All such possibilities have *prima facie* plausibility. But what about the computer vendors' promises we have all heard, promises that transformed millions of dollars into years of frustration? Have we not learned that the fancy new software packages promised 'in the fourth quarter' usually arrive, incomplete, with bugs and without documentation, a year late?

If we must place our bets, and we must, how can we distinguish the good bets from the bad bets? The answer to this question is difficult in the fairly prosaic fields of record keeping and manufacturing automation; it becomes even more elusive in the esoteric applications that AI addresses.

There are no conclusive arguments about the existence or non-existence of ultimate limits to the intelligence of machines, incompleteness theorems and uncertainty principles notwithstanding. Undoubtedly, however, some forms of intelligence are going to be easier to produce than others. AI scientists are still unable to assess realistically the relative difficulty of tasks. Some researchers, in their eagerness to convince us that automation has no ultimate limits not shared by human–biological–physical systems, still underestimate the distance between the possible and the imminent. We fail to distinguish the tasks that are well within reach from those we have not even got the foggiest notion of how to approach.

I here propose some rules of thumb (heuristics, in AI jargon) that may provide a first cut at separating the viable predictions from the wild claims. These heuristics are derived from my estimates of theoretical obstacles, and do not reflect such other factors as social resistance. For the moment I follow the custom of not attempting to justify heuristics, but simply present them as 'compiled hindsight', that is, generalization based on experience.

The first heuristic is this: the larger the proportion of the human population that can perform a particular activity, the less likely it is that the activity will soon be accomplished by computer. Expertise in structural engineering or medical diagnosis, soon; understanding a television sitcom, much later. Making a new discovery in the biology of reproduction, eventually; enjoying reproduction, maybe never.

However, this heuristic obviously fails in some cases; for example, more people can read a good novel than can write a good novel. In this case I believe it is safe to say that reading with understanding, though a skill of broader distribution, will be within the ability of computer software sooner than the skill of narrower distribution, writing creatively. My second heuristic thus adds a second dimension to the classification: analytic skills will fall to automation sooner than synthetic, productive skills. Reading before writing, proof checking before proof production, program testing before programming, plan execution before planning.

Intelligence is often conceived of as something purely cognitive, devoid of any action (motor) component. This suggests to some that motor skills are somehow secondary and hence less challenging for AI. My third heuristic makes just the opposite claim, that thought is often easier than action. Of course, automata that require human-level motor skills and problem-solving skills will be more difficult to produce than pre-programmed robots with limited movement. What my third heuristic claims is that the perceptual motor coordination and delicate controlled movements routinely achievable by every human arm will prove harder to simulate than much of the problem-solving skill that gets most of humanity through an ordinary day. Planning a menu with intelligence is well within reach of AI today; walking to the store to buy the food is not. Expert level chess strategy by machine is here today; replacing a captured piece with one's own in one smooth movement will elude us for some time. Intelligent planning of the family budget will be easier than making the beds.

Yet a fourth heuristic is needed: Both cognition and motor skills will be simulated before affect. This observation scarcely needs justification. Although I personally believe that even such pieces of the human experience as emotion, intentions, and consciousness will ultimately be explained by mechanistic principles, most people probably do not. In any case, most AI research ignores these issues, and there are very few serious attempts being made to address them directly from within the AI framework. The conventional wisdom within the research community is that such characteristics are emergent properties, that as systems become more complex and extensive the behavior of those systems will eventually be seen as possessing those properties. I do not hold to this view; but in any case, most would agree that AI is not hot on the trail of developing sentient, emotional beings.

Applying these four rules is not always straightforward, so let me provide some examples. I would suggest that the following capabilities will emerge in the relatively short run. (1) A proliferation of expert systems whose expertise is in relatively esoteric, atheoretic domains or in classical, well-formalized, narrow specialties. (2) Text processing and document preparation software that acts as a symbiotic checker of human produced writing, performing such tasks as stylistic checking, grammatical checking, and checking of the form of arguments, primarily in the domain of technical manuals, later technical writing more generally. (3) Teaching machines that shape human skills by automatic criticism and analysis of errors, with appropriate, immediate feedback. (4) Quality inspection systems employing visual, auditory, and tactile examination and limited manipulation of standardized items. (5) Checking of mathematical proofs presented in a formal calculus. (6) Automatic testing of computer software. (7) Alerting services that scan the print media for key ideas that may be expressed in subtle and varied ways. (8) Symbiotic couplings of man and machine, wherein the machines augment the strength and perceptual acuity of the human, while the human retains control of the planning and problem solving aspects of the task.

Some examples of programs not likely to be seen in the near future are the following (I recommend a critical stance toward programs that *appear* to have these properties). (1) Programs that can discuss in natural language a wide range of general, non-technical topics. (2) Knowledge based systems that know about the human condition, that is about those experiences not actually shared by the computer (travel, family life, fictional literature, television, religion, painting, music, athletic competition, birth, marriage, death and so forth). (3) Decision making systems that permit a manager to exercise control over large groups of other humans merely by interacting with a computer, without face-to-face hard bargaining. (4) Any computer system that does significant important planning without human intervention at the highest level. (5) Programs that are the source of fundamental new ideas in science, technology, or the arts. (6) Programs that depend on knowledge of human culture and cultural differences, e.g., expert systems in diplomacy and foreign relations.

Having attempted to list some examples of extreme cases, let me address the nature of the theoretical difficulty underlying my heuristics. Psychology, and philosophy before it, traditionally divides human mental life into categories, including cognition, perception, motivation, emotion, and motor control. AI subscribes to the working hypothesis that all aspects of mental activity can be explained by a computational model. However, a second major AI premise is that there are purely cognitive activities that can be considered separately, and that a broad and important range of human productivity will be encompassed thereby. 'Intelligence' carries this narrowed connotation, and it is significant that AI is not AU (artificial understanding), AT (artificial thought), or AM (artificial mind). Similarly, robotics (artificial action) has usually been considered a separate discipline. It is no doubt true that in this narrowed sense, machines have exhibited intelligence. But the reason that AI has often bitten off more than it can chew is that most of human activity, even 'purely intellectual' activity, does not consist of isolated acts of cognition, unconnected to human experience. The interactions are often subtle, and, unfortunately, easily masked by careful selection of examples. Natural language understanding is perhaps the best illustration of this. It is easy to believe, even compelling, that a program that replies in a natural language has understood what it was told. However, a skillful and skeptical examination of a program's behavior soon unmasks it.

The thrust of my suggestions is that putative skills that involve, however subtly, affect, consciousness, and experience in the real world of physical and social activity will not soon be achieved by computer; that is, not until substantive new ideas come forth to address these aspects of mental life at least as well as the idea of heuristic programming addresses cognition.

Armed with these heuristics and the examples cited, perhaps a manager or planner will be in a better position to separate the hyperbole from the plausible. Of course these are simply my opinions, and as such just another data point to fold into the forecasting formula. Several other opinions and many more examples, covering a wide range, are presented in this special issue of *Human Systems Management* devoted to artificial intelligence. I have chosen to make my introductory comments conservative rather than evangelical in order to put the reader in a critical mood.

The papers, and the shorter commentaries, occupy a long stretch of the forecasters' limb. Each is written by respected authors, contributors, and observers of the AI scene, and each deserves careful consideration. I hope this issue will be of substantial interest and application to our readers.